Statistics

Lecture 4 & 5

Random variable

SILESIAN UNIVERSITY
SCHOOL OF BUSINESS ADMINISTRATION IN KARVINA

David Bartl
Statistics
INM/BASTA

Outline of the lecture

- The concept of probability
- Random variable
- Measures of central tendency (mean, mode, median)
- Measures of dispersion (variance)
- Measures of shape (skewness, kurtosis)
- Functions of random variables (sample mean, sample variance)

The concept of probability

An **experiment** is any physical procedure which can end up with a result from a set of possible outcomes, and the experiment can be repeated (up to) infinitely many times.

The final result ω of an experiment is called an **outcome** and the set Ω of all the outcomes is called the **sample space**.

An **event** E is a set of outcomes, i.e. a subset of the sample space $(E \subseteq \Omega)$.

An **elementary event** is the singleton $E = \{\omega\}$ for any outcome $\omega \in \Omega$.

The **impossible event** is the empty set $E = \emptyset$.

The **certain event** is the sample space $E = \Omega$ itself.

The concept of probability

The **event space** $\mathcal{F}$ is the collection of all (measurable) events. The event space $\mathcal{F}$ is a σ-algebra of the subsets of the sample space Ω (i.e. $\Omega \in \mathcal{F}$ and, for any $E, E_1, E_2, E_3, \ldots \in \mathcal{F}$, we have $\Omega \setminus E \in \mathcal{F}$ and $\bigcup_{i=1}^{\infty} E_i \in \mathcal{F}$).

The event space contains all subsets of the sample space (i.e. $\mathcal{F} = \{E : E \subseteq \Omega\}$) especially in discrete cases, i.e. when the sample space is finite ($\Omega = \{1, 2, \ldots, N\}$, e.g., when tossing a coin, rolling a dice, etc.) or countably infinite ($\Omega = \{1, 2, 3, \ldots\}$).

The concept of probability

The event space does not contain all subsets of the sample space (i.e. $\mathcal{F} \neq \{E : E \subseteq \Omega\}$) especially in continuous cases, when the probability is defined in the geometrical way or when dealing with many standard continuous probability distributions (normal, exponential, uniform, etc.), i.e. when the probability is connected with the Lebesgue measure on the real numbers $\mathbb{R}$.

It holds in general only that

$$\emptyset \subset \mathcal{F} \subseteq \{E : E \subseteq \Omega\}$$

When considering an event E, we shall always mean an event such that $E \in \mathcal{F}$.

The concept of probability

By Kolmogorov's definition, the probability is a function $P: \mathcal{F} \to \mathbb{R}$ such that $P(\Omega) = 1$ and, for any pairwise disjoint events $E, E_1, E_2, E_3, \ldots \in \mathcal{F}$, we have $P(E) \geq 0$ and $P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$.

Despite this elegant definition, which allows us to grasp the concept of the probability mathematically, there are several interpretations what the probability actually is:

- classical definition ("all the elementary events are 'equally likely' to occur")
- frequentist definition (repeat the random experiment infinitely many times)
- Bayesian probability

Random variable

Consider a probability space, i.e. a triple $(\Omega, \mathcal{F}, P)$ where

- Ω **is the sample space** (the set of all possible outcomes of a random experiment)
- $\mathcal{F}$ **is the event space** (the collection of all measurable events)
- P **is the probability** (a non-negative σ-additive function $P: \mathcal{F} \to \mathbb{R}$ such that $P(\Omega) = 1$)

A **random variable** is any function

$$X: \Omega \longrightarrow \mathbb{R}$$

which is measurable, i.e. the preimage of any open interval is an event ($X^{-1}\big((a, b)\big) = \{\omega \in \Omega : X(\omega) \in (a, b)\} \in \mathcal{F}$ for every $a, b \in \mathbb{R}$ such that $a < b$).

Random variable

Since the sample space Ω is the set of all possible outcomes of a random experiment and the random variable X is a (measurable) function $X: \Omega \to \mathbb{R}$, the random variable can be seen as the numerical outcome of the random experiment.

<u>Notice:</u> Any quantitative (numerical) <u>data item</u> of the data units of the dataset can be seen as a <u>random variable</u>.

Random variable

Let a probability space $(\Omega, \mathcal{F}, P)$ be given. Since the random variable X is a (measurable) mapping $X: \Omega \to \mathbb{R}$, it follows that the distribution of the (numerical) values of the variable X is governed by the probability P.

To simplify the matters significantly, we shall **assume** from now on that

there exists a probability density function of the probability measure P.

Assumptions to simplify the matters

We shall assume in particular that exactly one the three case arises:

I. the sample space Ω is finite

II. the sample space Ω is countably infinite

III. the sample space $\Omega = \mathbb{R}$

Each of the cases is discussed in detail below.

Assumptions to simplify the matters

<u>Case I</u>: The sample space Ω is finite, such as

$\Omega = \{\text{heads}, \text{tails}\}$ (when tossing a coin)

$\Omega = \{\text{up}, \text{down}\}$ (when tossing a tack)

$\Omega = \{1, 2, 3, 4, 5, 6\}$ (when rolling a dice)

Etc.

We may assume in general that

$$\Omega = \{1, 2, \ldots, N\}$$

where N is the number of elements of the sample space Ω.

Assumptions to simplify the matters

<u>Case II</u>: The sample space Ω is countably infinite, such as

$$\Omega = \{1, 2, 3, 4, 5, \ldots\}$$

$$\Omega = \{0, 1, 2, 3, 4, 5, \ldots\}$$

$$\Omega = \{\ldots, -3, -2, -1, 0, +1, +2, +3, \ldots\}$$

Etc.

To illustrate the second example $(\Omega = \{0, 1, 2, \ldots\})$,

consider the next random experiment:

Tossing a coin, count the number of “heads” until the first “tails” occur.

Assumptions to simplify the matters

Cases I & II: If the sample space is finite $(\Omega = \{1, 2, \ldots, N\})$ or countable $(\Omega = \{1, 2, 3, \ldots\})$ we assume that the event space $\mathcal{F}$ is the collection of all subsets of the sample space, i.e. $\mathcal{F} = 2^{\Omega}$, i.e.

$$\mathcal{F} = \{E : E \subseteq \Omega\}$$

Then there exists a **probability mass function** of the given probability P.

Probability mass function

Given the probability space $(\Omega, \mathcal{F}, P)$, the **probability mass function** of the probability measure P is a function

$$p: \Omega \rightarrow \mathbb{R}$$

such that it holds

$$P(E) = \sum_{\omega \in E} p(\omega) \qquad \text{for every event} \quad E \in \mathcal{F}$$

Probability mass function

In the cases I & II, when the sample space Ω is finite $(\Omega = \{1, 2, \dots, N\})$ or countable $(\Omega = \{1, 2, 3, \dots\})$, and $\mathcal{F} = 2^{\Omega}$, the probability mass function $p: \Omega \to \mathbb{R}$ clearly exists. It is enough to put

$$p(\omega) = P(\{\omega\}) \qquad \text{for every} \quad \omega \in \Omega$$

Then

$$P(E) = \sum_{\omega \in E} p(\omega) \qquad \text{for every event} \quad E \in \mathcal{F}$$

Notice also that

$$p(\omega) \in [0, 1] \qquad \text{for every} \quad \omega \in \Omega$$

Assumptions to simplify the matters

In the cases I & II, when the sample space Ω is finite or countable and $\mathcal{F} = 2^{\Omega}$, the probability mass function p is also the density function of the probability P with respect to the counting measure.

The counting measure α is a function such that $\alpha(E) = |E|$, the number of elements of the set E if the set E is finite, or $\alpha(E) = +\infty$ if the set E is infinite. Then

$$P(E) = \sum_{\omega \in E} p(\omega) = \int_E p(\omega)\,\mathrm{d}\alpha$$

Assumptions to simplify the matters

Recall that, in the cases I & II, when the sample space Ω is finite or countable, and $\mathcal{F} = 2^{\Omega}$, we assume that the random variable X is any function

$$X: \Omega \to \mathbb{R}$$

<u>Case III</u>: We assume that

- the sample space $\Omega = \mathbb{R}$ is the set of the real numbers
- the event space $\mathcal{F}$ is the collection of all Lebesgue measurable subsets of $\mathbb{R}$
- the random variable X is the identity function

$$X: \mathbb{R} \to \mathbb{R} \qquad X(x) = x \qquad \text{for every} \quad x \in \mathbb{R}$$

Probability density function

Given the probability space $(\Omega, \mathcal{F}, P)$, the **probability density function** of the probability measure P with respect to a reference measure λ on $(\Omega, \mathcal{F})$ is a (measurable) function

$$f: \Omega \to \mathbb{R}$$

such that it holds

$$P(E) = \int_E f(\omega)\, \mathrm{d}\lambda \qquad \text{for every event} \quad E \in \mathcal{F}$$

(The integral on the right-hand side is the Lebesgue integral.)

Lebesgue measure and Lebesgue integral

It is beyond the scope of this lecture to construct the Lebesgue measure λ on $\mathbb{R}$, the collection of the Lebesgue measurable sets, and to introduce the Lebesgue integral. That is why we shall work with the integral "intuitively".

Indeed, **if the density function** $f: \mathbb{R} \to \mathbb{R}$ **is continuous** and the event $E \in \mathcal{F}$ is an **interval** $E = (a, b), [a, b), (a, b], [a, b]$ (with $-\infty \leq a \leq b \leq +\infty$, but $\pm\infty \notin E$), **then**

$$P(E) = \int_E f(x)\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x$$

Lebesgue measure and Lebesgue integral

Actually, the latter case (the event E is an interval and the density function f is continuous) is the only case which we shall need in practice.

By seeing the Kolmogorov theory of probability as a special case of the measure theory, we could treat all the cases I, II, and III in a uniform way (together at once).

For "simplicity" (because the measure theory is beyond the scope of this lecture), however, we treat the cases I & II and the case III separately.

Assumptions to simplify the matters

<u>Cases I & II</u>: We assume for simplicity that

- the sample space Ω is finite or countable (such as $\Omega = \{1, \ldots, N\}$ or $\Omega = \{1, 2, \ldots\}$)
- the event space $\mathcal{F} = 2^{\Omega} = \{E : E \subseteq \Omega\}$ is the collection of all subsets of Ω
- the random variable X is any function

$$X: \Omega \to \mathbb{R}$$

$$X: \omega \mapsto X(\omega) \qquad \text{for every} \quad \omega \in \Omega$$

- and there exists a probability mass function $p: \Omega \to \mathbb{R}$ such that

$$P(E) = \sum_{\omega \in E} p(\omega) \qquad \text{for every event} \quad E \in \mathcal{F}$$

Assumptions to simplify the matters

Case III: We assume for simplicity that

- the sample space $\Omega = \mathbb{R}$ is the set of the real numbers
- the event space $\mathcal{F}$ is the collection of all Lebesgue measurable subsets of $\mathbb{R}$
- the random variable X is the identity function

$$X: \mathbb{R} \to \mathbb{R}$$

$$X: x \mapsto x \qquad \text{for every} \quad x \in \mathbb{R}$$

- and there exists a continuous probability density function $f: \mathbb{R} \to \mathbb{R}$ such that

$$P(E) = \int_E f(x)\,\mathrm{d}x \qquad \text{for every event} \quad E \in \mathcal{F}$$

Examples of random variables

There are 100 rooms in some hotel.

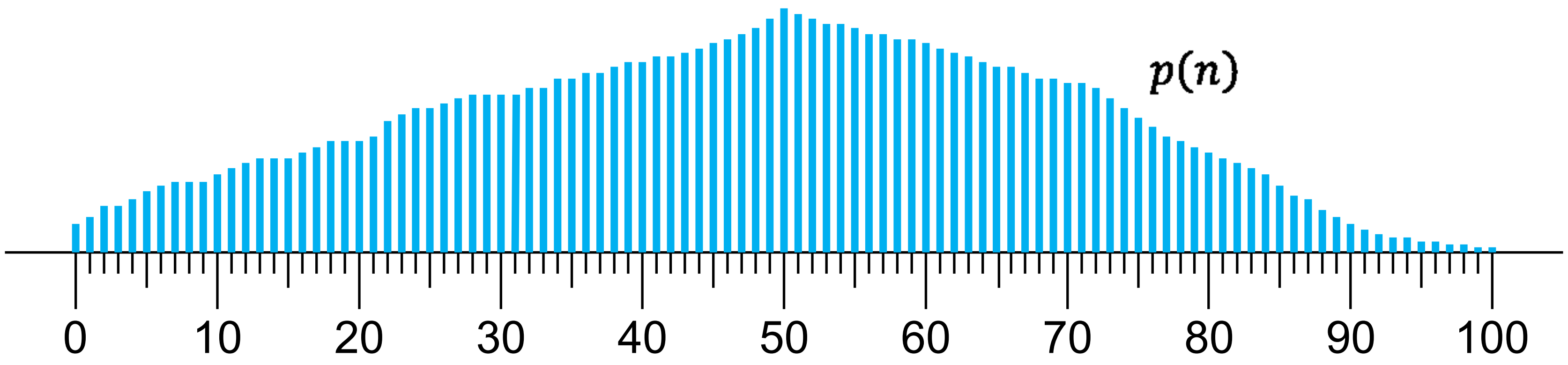
The number of the occupied rooms is a random variable X.

The possible values of this random variable are numbers 0, 1, 2, …, 100.

In other words, the range of the random variable is the set {0, 1, 2, …, 100}.

The **probability mass function** of the random variable X may look, e.g.,

like this:



$$\sum_{n=0}^{100} p(n) = 1$$

Examples of random variables

A company's employees' salary per year can be seen as a random variable X.

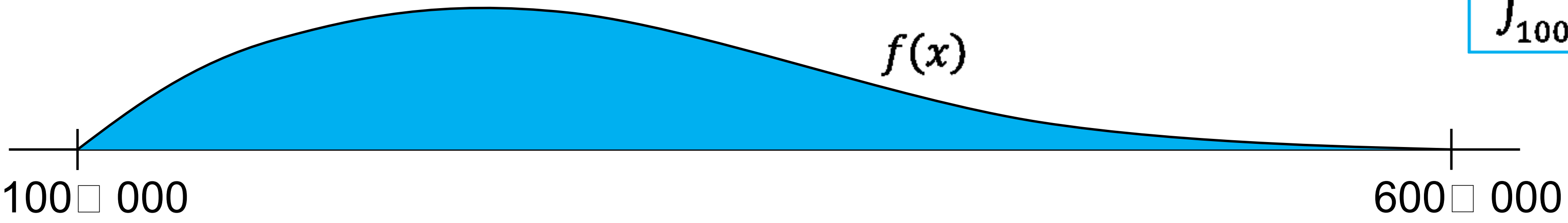
The salary per year is in the range from 100 000 to 600 000 of monetary units.

Seeing the salary as a continuous random variable, then

the **probability density function** of the random variable X may look, e.g.,

like this:

$$\int_{100\,000}^{600\,000} f(x)\,\mathrm{d}x = 1$$

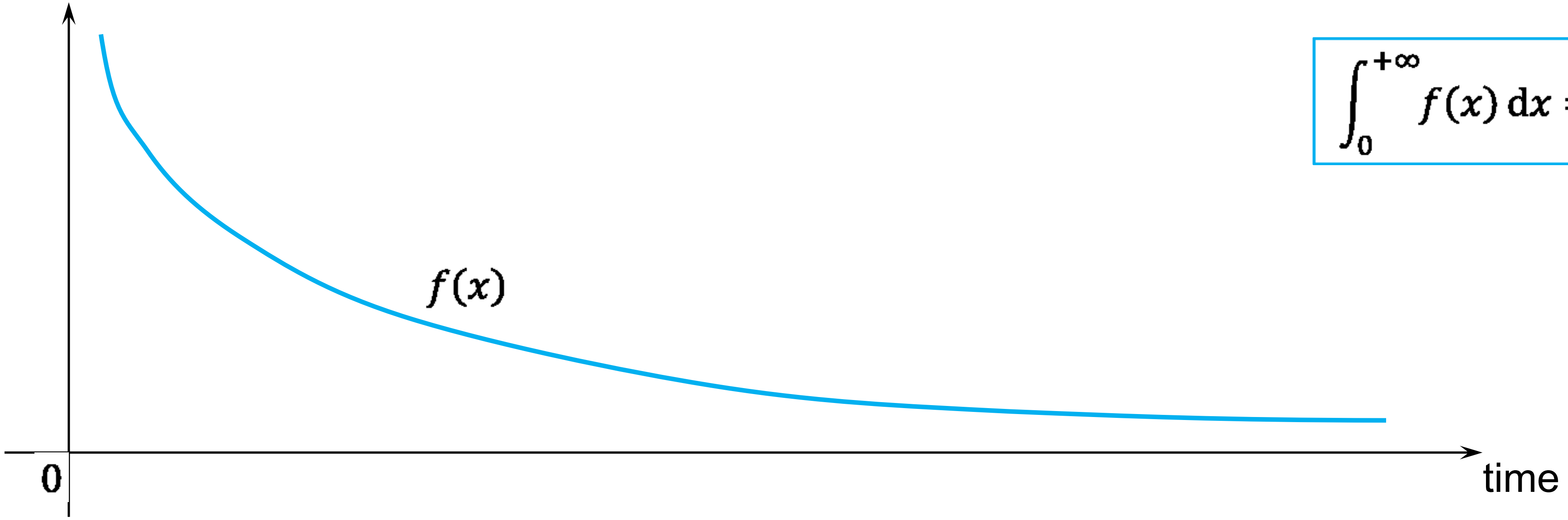


Examples of random variables

The lifetime of a product (such as a bulb) is a continuous random variable X.

This random variable can attain any non-negative value.

The **probability density function** of the random variable X may look, e.g., like this:



$$\int_0^{+\infty} f(x)\,\mathrm{d}x = 1$$

Examples of random variables

The “wheel of fortune”.

The customer rotates the wheel and, depending upon the final position, the discount of the price is deduced.

Examples of random variables

Example: The "wheel of fortune".

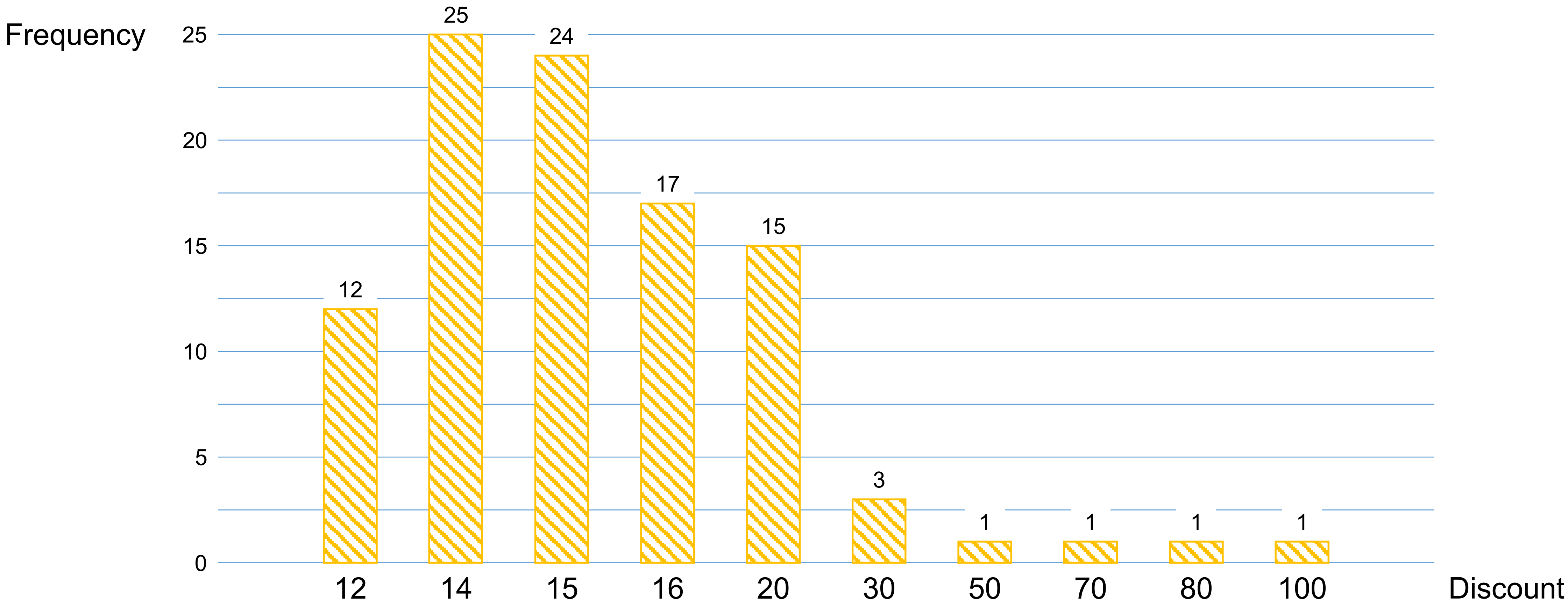
The sample space: $\Omega = \{A, B, C, D, E, F, G, H, I, J\}$

Table:

$\omega \in \Omega$	$x_\omega = X(\omega)$ Discount in %	n_ω Frequency	$p_\omega = p(\omega)$ Relative frequency
A	□ 12	□ 12	□ 12 %
B	□ 14	□ 25	□ 25 %
C	□ 15	□ 24	□ 24 %
D	□ 16	□ 17	□ 17 %
E	□ 20	□ 15	□ 15 %
F	□ 30	□ □ 3	□ □ 3 %
G	□ 50	□ □ 1	□ □ 1 %
H	□ 70	□ □ 1	□ □ 1 %
I	□ 80	□ □ 1	□ □ 1 %
J	100	□ □ 1	□ □ 1 %
	TOTAL	100	100 %

Examples of random variables

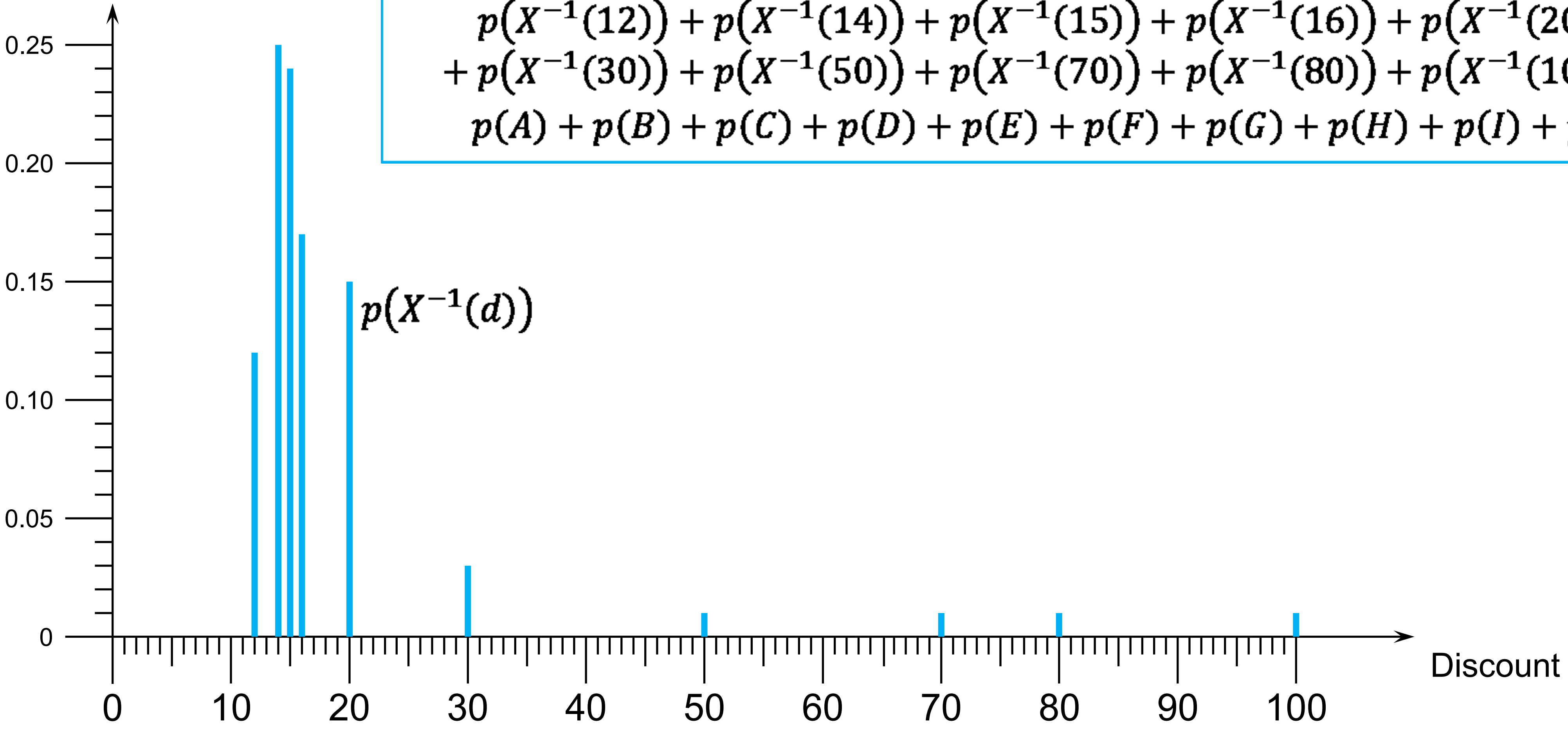
Bar chart – the frequencies (numbers) of the ordinal data item "Discount":



Examples of random variables

The graph of the probability mass function of the random variable X:

$$p\left(X^{-1}(12)\right)+p\left(X^{-1}(14)\right)+p\left(X^{-1}(15)\right)+p\left(X^{-1}(16)\right)+p\left(X^{-1}(20)\right)+$$
$$+p\left(X^{-1}(30)\right)+p\left(X^{-1}(50)\right)+p\left(X^{-1}(70)\right)+p\left(X^{-1}(80)\right)+p\left(X^{-1}(100)\right)=$$
$$p(A)+p(B)+p(C)+p(D)+p(E)+p(F)+p(G)+p(H)+p(I)+p(J)=1$$



Probability mass function & Probability density function

<u>Cases I & II</u>: Recall that a **probability mass function** is any function $p\colon \Omega \to \mathbb{R}$

such that

$$\sum_{\omega \in \Omega} p(\omega) = 1 \qquad \text{and} \qquad p(\omega) \geq 0 \qquad \text{for every} \quad \omega \in \Omega$$

<u>Case III</u>: Recall that a **probability density function** is any function $f\colon \mathbb{R} \to \mathbb{R}$

such that

$$\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = 1 \qquad \text{and} \qquad f(x) \geq 0 \qquad \text{for every} \quad x \in \mathbb{R}$$

Cumulative distribution function

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X: \Omega \to \mathbb{R}$ be a random variable.

Then the **cumulative distribution function** of the random variable X

is the function

$$F: \mathbb{R} \to \mathbb{R}$$

defined by

$$F(x) = P(\{\omega \in \Omega : X(\omega) \leq x\})$$

Notice that the expression "$P(\{\omega \in \Omega : X(\omega) \leq x\})$" is often written as "$P(X \leq x)$" for short.

Cumulative distribution function

The cumulative distribution function $F(x) = P(X \leq x)$ is

- non-decreasing
- right-continuous

and it also holds

$$\lim_{x \to -\infty} F(x) = 0 \qquad \text{and} \qquad \lim_{x \to +\infty} F(x) = 1$$

Moreover, any function $F: \mathbb{R} \to \mathbb{R}$ that satisfies the above properties is the cumulative distribution function of some random variable.

Cumulative distribution function

By the definition

$$F(x) = P(X \leq x)$$
$$= P(\{\omega \in \Omega : X(\omega) \leq x\})$$

of the cumulative distribution function, it follows that

$$P(a < X \leq b) =$$
$$P(\{\omega \in \Omega : a < X(\omega) \leq b\}) = F(b) - F(a) \qquad \text{for every} \quad a, b \in \mathbb{R}$$

Examples of cumulative distribution functions

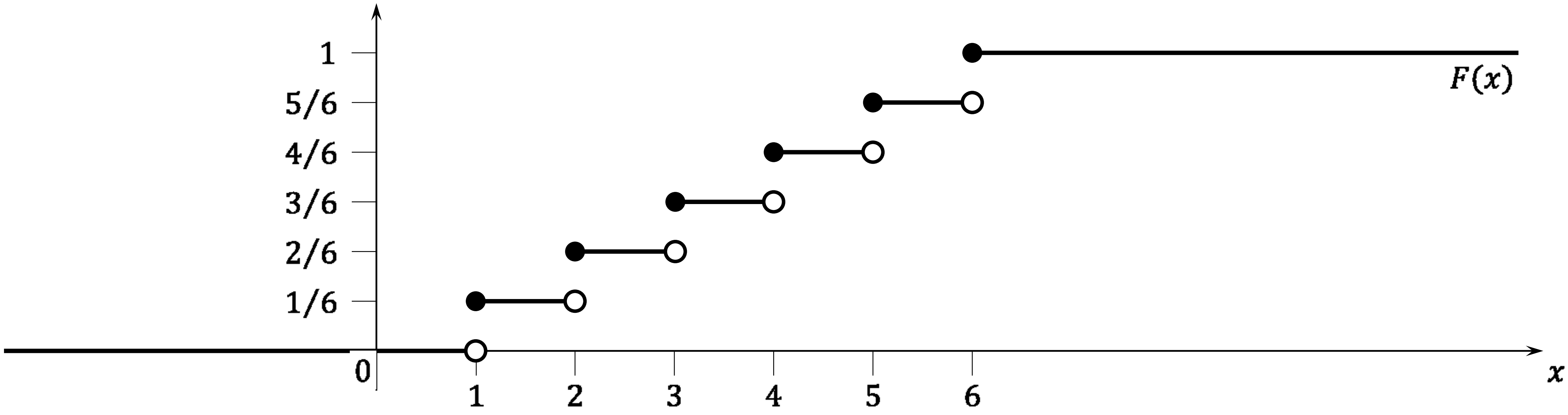
Example: Rolling a dice.

The random variable: $X(x) = x$

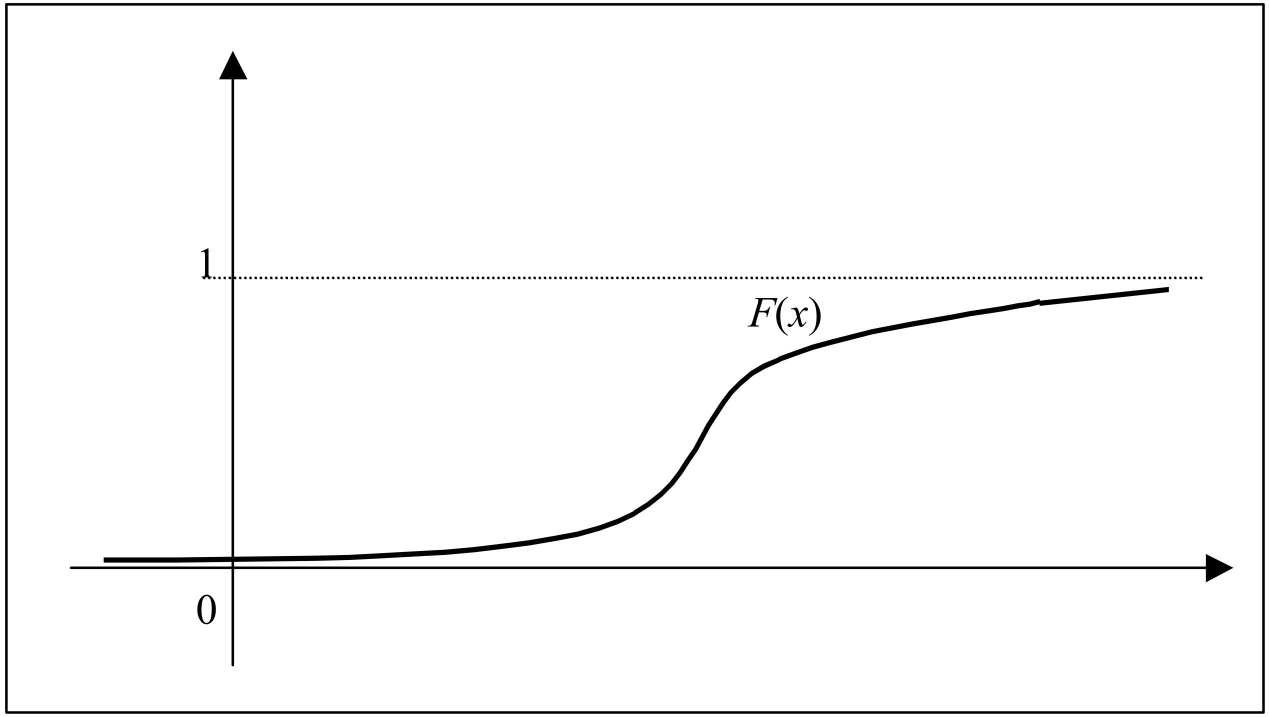
The sample space: $\Omega = \{1, 2, 3, 4, 5, 6\}$

The probability mass function: $p(x) = \frac{1}{6}$

The **graph** of the **cumulative distribution function** of the random variable X:



Examples of cumulative distribution functions



An example of a cumulative distribution function of a continuous random variable.

Recall that F is non-decreasing,

$$\lim_{x\to-\infty} F(x) = 0$$

and

$$\lim_{x\to+\infty} F(x) = 1$$

The cumulative distribution & the density function

Case III: Let $(\Omega, \mathcal{F}, P)$ be a probability space where the sample space $\Omega = \mathbb{R}$, the event space $\mathcal{F}$ is the collection of all Lebesgue measurable subsets of $\mathbb{R}$, and the probability P is such that there exists a continuous probability density function f such that $P(E) = \int_E f(x)\,\mathrm{d}x$ for every event $E \in \mathcal{F}$.

Consider the identity random variable $X(x) = x$ for every $x \in \mathbb{R}$.

Then

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t \qquad \text{and} \qquad f(x) = F'(x) \qquad \text{for every} \quad x \in \mathbb{R}$$

The cumulative distribution & the density function

Case III: Let $(\Omega, \mathcal{F}, P)$ be a probability space where the sample space $\Omega = \mathbb{R}$, the event space $\mathcal{F}$ is the collection of all Lebesgue measurable subsets of $\mathbb{R}$, and the probability P is such that there exists a continuous probability density function f such that $P(E) = \int_E f(x)\,\mathrm{d}x$ for every event $E \in \mathcal{F}$.

Consider the identity random variable $X(x) = x$ for every $x \in \mathbb{R}$.

It holds by definitions

$$F(x) = P(X \leq x) = P(\{t \in \mathbb{R} : -\infty < t \leq x\}) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$$

By the fundamental theorem of calculus,

$$f(x) = F'(x) \qquad \text{for every} \quad x \in \mathbb{R}$$

The cumulative distribution & the density function

<u>Case III</u>: Let $(\Omega, \mathcal{F}, P)$ be a probability space where the sample space $\Omega = \mathbb{R}$, the event space $\mathcal{F}$ is the collection of all Lebesgue measurable subsets of $\mathbb{R}$, and the probability P is such that there exists a <u>continuous</u> probability density function f such that $P(E) = \int_E f(x)\,\mathrm{d}x$ for every event $E \in \mathcal{F}$.

Consider the identity random variable $X(x) = x$ for every $x \in \mathbb{R}$.

If $-\infty < a \leq b < +\infty$, then it holds by the continuity of the density function f that

$$\int_{(a,b)} f(x)\,\mathrm{d}x = \int_{[a,b)} f(x)\,\mathrm{d}x = \int_{(a,b]} f(x)\,\mathrm{d}x = \int_{[a,b]} f(x)\,\mathrm{d}x$$

hence

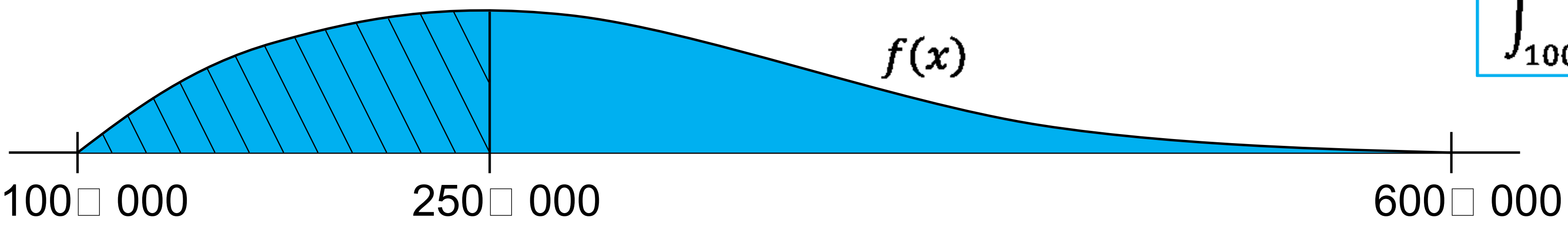
$$P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b) = \int_a^b f(x)\,\mathrm{d}x$$

The probability of an event

Let the **probability density function** of a random variable X

(the employees' salary per year) look like this:

$$\int_{100\,000}^{600\,000} f(x)\,\mathrm{d}x = 1$$



Then the probability the salary of an employee is in the range from 100000 to 250 000, say, is:

$$P(100\,000 \leq X \leq 250\,000) =$$

$$= P(\{x \in [100\,000, 600\,000] : 100\,000 \leq X(x) \leq 250\,000\}) = \int_{100\,000}^{250\,000} f(x)\,\mathrm{d}x$$

Measures of central tendency

- Mean / Expected value
- Mode
- Quantile
- Median
- Quartiles
- Deciles
- Centiles

The mean / expected value

The **expected value** (or the mean value) of the random variable X is denoted by

$$\mu \quad \text{or} \quad \mathrm{E}[X]$$

Cases I & II:

$$\mu = \mathrm{E}[X] = \sum_{\omega \in \Omega} X(\omega) p(\omega)$$

Case III:

$$\mu = \mathrm{E}[X] = \int_{-\infty}^{+\infty} x f(x) \, \mathrm{d}x$$

The mode

Roughly speaking, the **mode** of the random variable X is the most probable value that the variable will attain.

Alternatively, the **mode** is the most frequent value of the random variable X (cf. the frequentist definition of the probability).

The definition of the mode is different for discrete variables (cases I & II) and for continuous variables (case III).

The mode

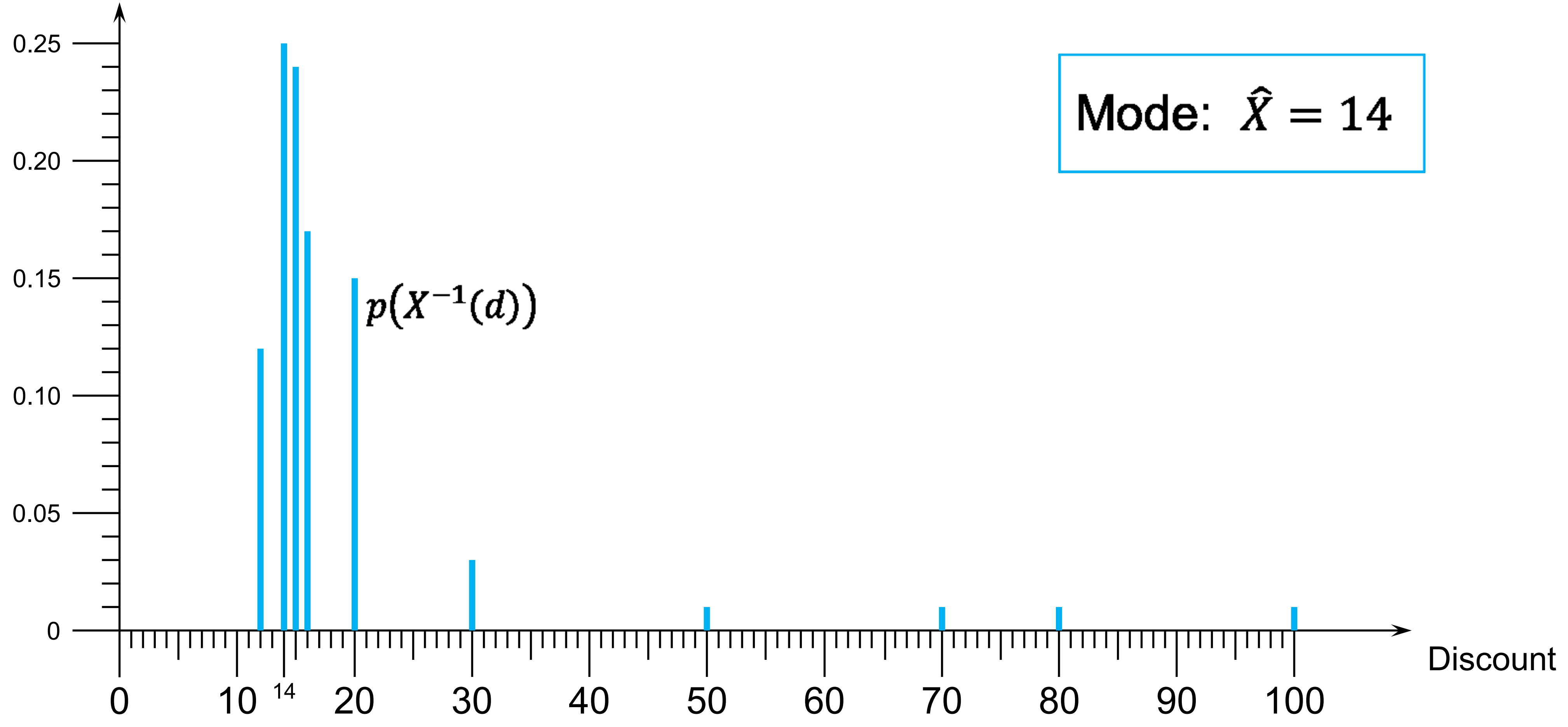
Cases I & II: Let $(\Omega, \mathcal{F}, P)$ be a probability space where the sample space Ω is finite or countably infinite, the event space $\mathcal{F} = 2^{\Omega}$, let p be the probability mass function of the probability P, and let $X\colon \Omega \to \mathbb{R}$ be a random variable.

The number $\hat{X} \in \mathbb{R}$ is a **mode** of the random variable X if and only if

$$\sum_{\substack{\omega \in \Omega \\ X(\omega)=\hat{X}}} p(\omega) \geq \sum_{\substack{\omega \in \Omega \\ X(\omega)=x}} p(\omega) \qquad \text{for every other} \quad x \in \mathbb{R}$$

The mode

The graph of the probability mass function of the random variable X:



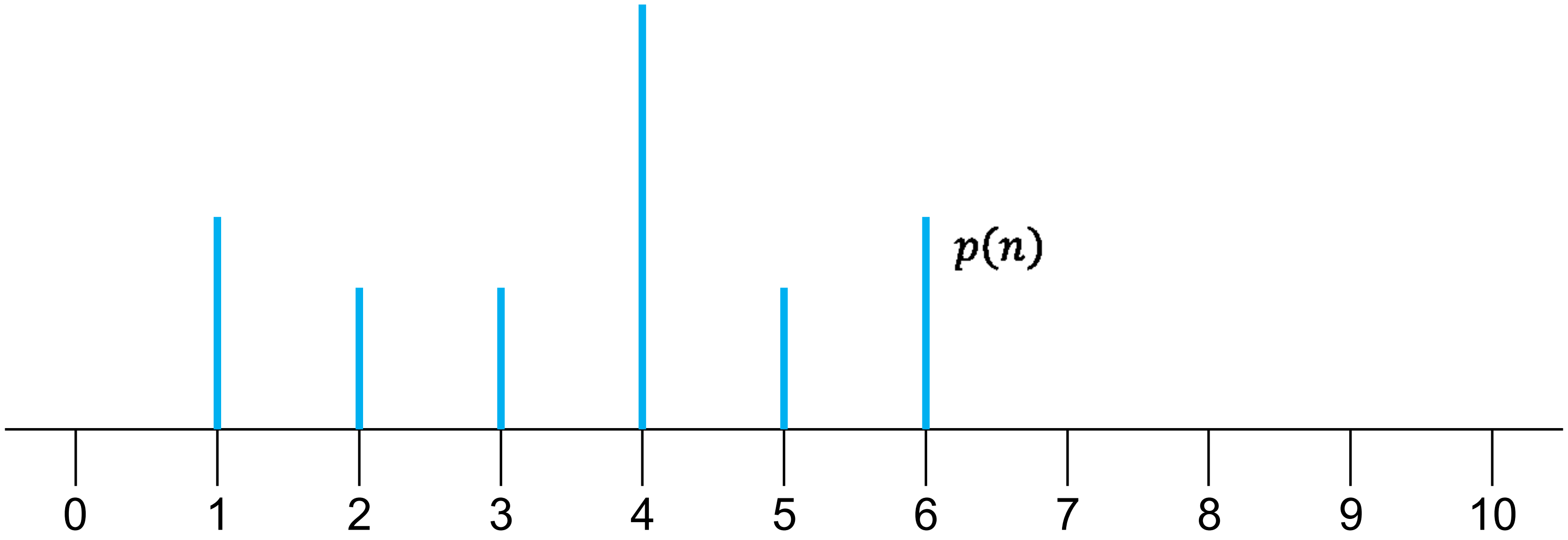
Mode: $\hat{X} = 14$

The mode

Example: Some example with some probability mass function.

The sample space: $\Omega = \{1, 2, 3, 4, 5, 6\}$ **The random variable:** $X(x) = x$

Mode: $\hat{X} = 4$

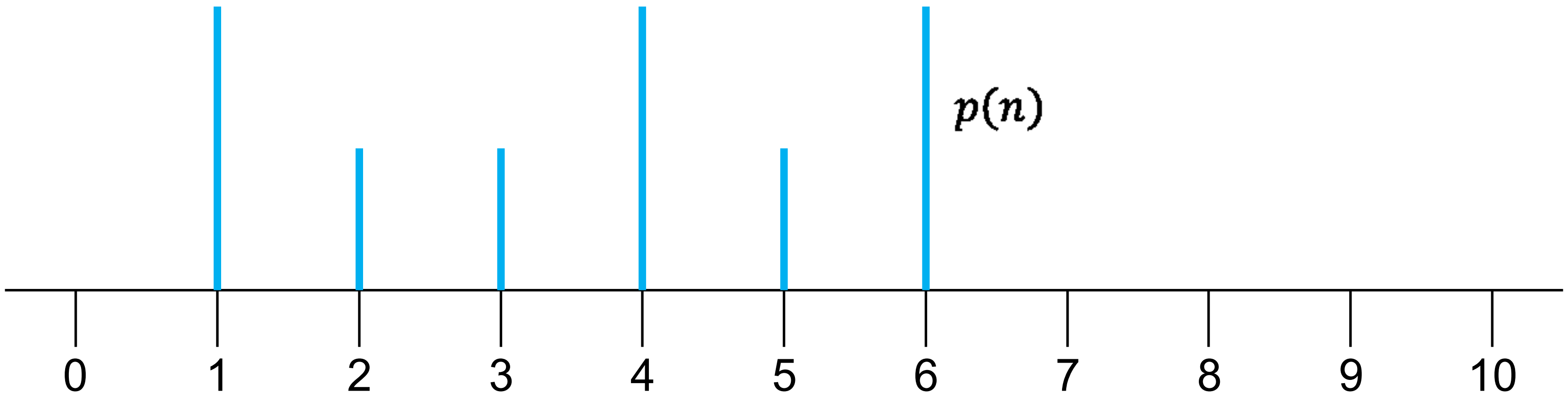


The mode

Example: Some example with some probability mass function.

The sample space: $\Omega = \{1, 2, 3, 4, 5, 6\}$ **The random variable:** $X(x) = x$

Modes: $\hat{X} = 1, 4, 6$



The mode

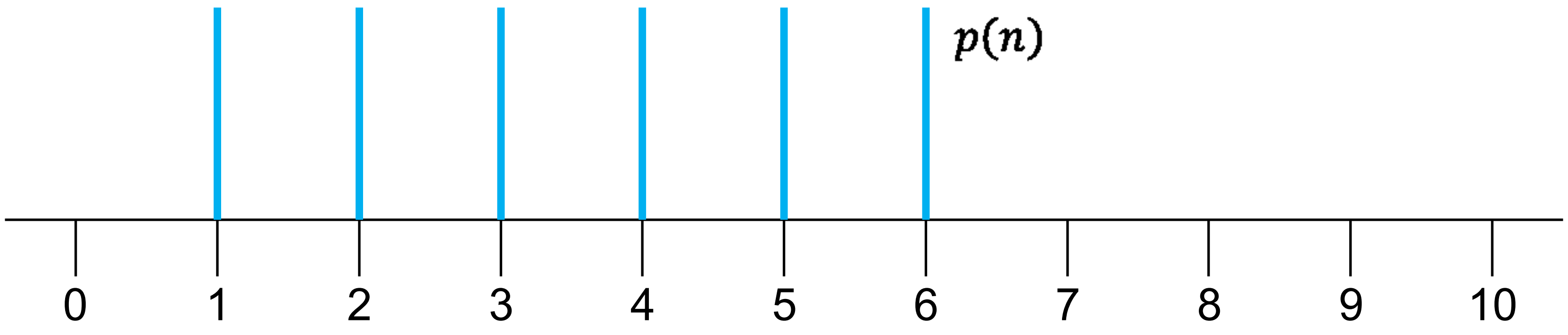
Example: Rolling a dice.

The sample space: $\Omega = \{1, 2, 3, 4, 5, 6\}$

The random variable: $X(x) = x$

The probability mass function: $p(x) = \frac{1}{6}$

Modes: $\hat{X} = 1, 2, 3, 4, 5, 6$



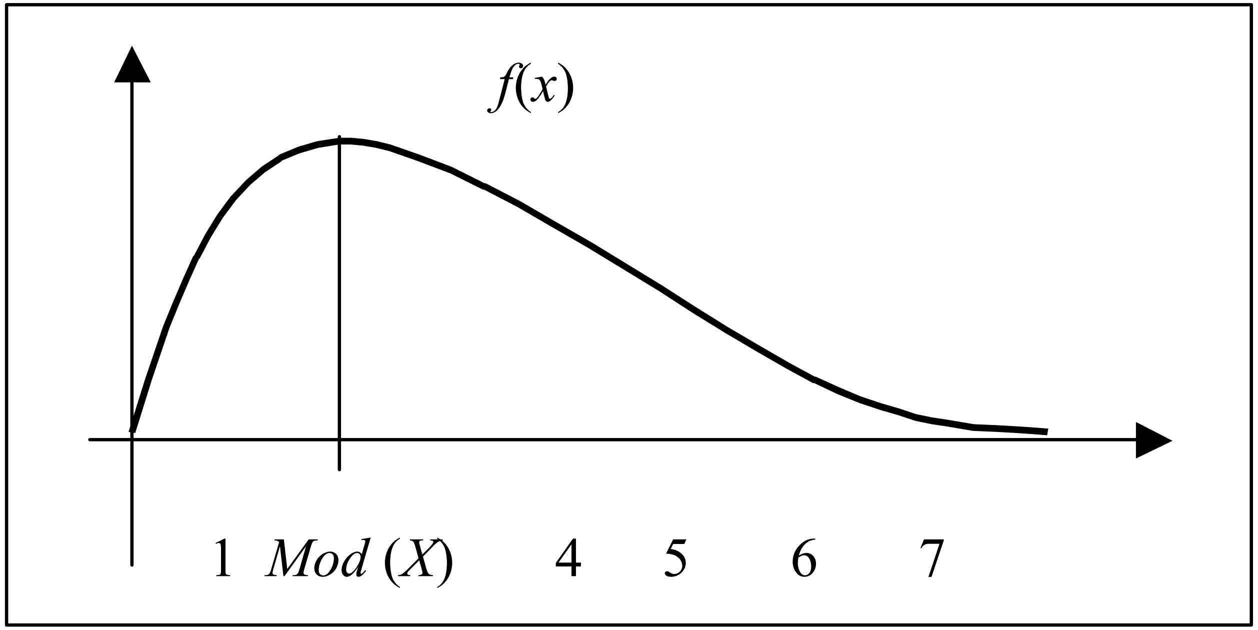
The mode

Case III: Let $(\Omega, \mathcal{F}, P)$ be a probability space where the sample space $\Omega = \mathbb{R}$, the event space $\mathcal{F}$ is the collection of all Lebesgue measurable subsets of $\mathbb{R}$, let f be the probability density function of the probability P, assume that f is continuous, and let $X(x) = x$ for every $x \in \mathbb{R}$.

The number $\hat{X} \in \mathbb{R}$ is a **mode** of the random variable X if and only if

there is a local maximum of the density function f at the point $\hat{X}$

The mode



The **mode** is any point at which the density function $f(x)$ attains its local maximum.

The mode

If the random variable is discrete (i.e. there is a probability mass function of the given probability measure P) or the random variable is continuous and there is a continuous probability density function of the given probability P, then there exists at least one mode of the random variable X.

In other words, under our assumptions (cases I & II or III), at least one mode of the random variable X exists.

The mode

Remarks:

- If the random variable is continuous, but the probability density function of the given probability P is not continuous, then the mode of the random variable may not exist.
- If the random variable is continuous and there exists no probability density function of the given probability P, then the mode of the random variable can not be defined at all !

The mode

Remarks:

- There may exist more than one mode.
- The probability distribution is termed:
  - **unimodal**, if there is exactly one mode
  - **bimodal**, if there are exactly two modes
  - etc.
  - **multimodal**, if there are several modes

Quantile

Let $(\Omega, \mathcal{F}, P)$ be any probability space and let $X: \Omega \to \mathbb{R}$ be any random variable. Then the **quantile** corresponding to a given probability $p \in [0,1]$ with respect to the cumulative distribution function $F(x) = P(X \leq x)$ of the random variable X is the value $x_p \in \mathbb{R}$ such that

$$P(X < x_p) \leq p \leq P(X \leq x_p)$$

Since the cumulative distribution function $F(x) = P(X \leq x)$ of the random variable X is right-continuous, it is equivalent to say that the **quantile** is the least value $x_p \in \mathbb{R}$ such that

$$p \leq P(X \leq x_p)$$

Quantile

Let $(\Omega, \mathcal{F}, P)$ be any probability space and let $X: \Omega \to \mathbb{R}$ be any random variable. If the cumulative distribution function $F(x) = P(X \leq x)$ of the random variable X is continuous and strictly monotonically increasing, then the **quantile** is the value $x_p \in \mathbb{R}$ such that

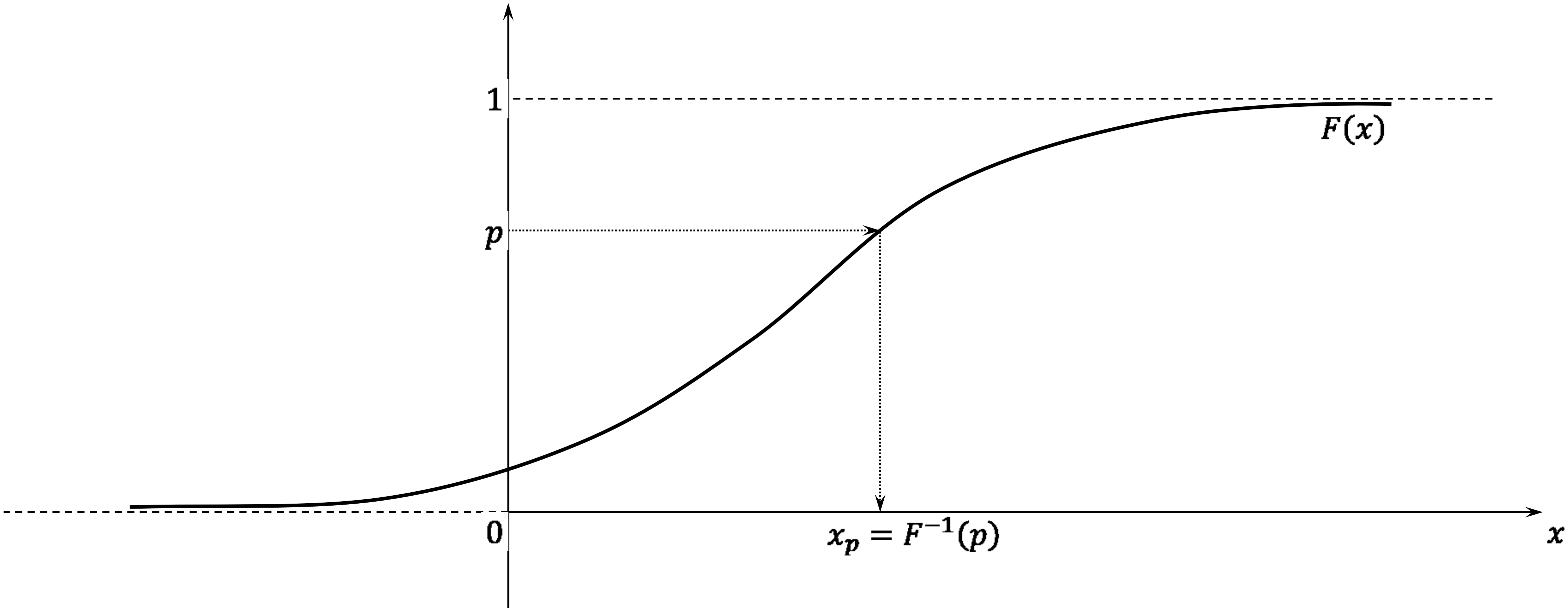
$$P(X \leq x_p) = p$$

i.e.

$$x_p = F^{-1}(p)$$

where F^{-1} is the function inverse to the cumulative distribution function F.

Quantile



Median

Let $(\Omega, \mathcal{F}, P)$ be any probability space and let $X: \Omega \to \mathbb{R}$ be any random variable. Then the **median** with respect to the cumulative distribution function $F(x) = P(X \leq x)$ of the random variable X is the **quantile corresponding to the probability** $p = 0.5$, i.e. the value $\tilde{X} = x_{0.5} \in \mathbb{R}$ such that

$$P(X < x_{0.5}) \leq \frac{1}{2} \leq P(X \leq x_{0.5})$$

Quartiles

Let $(\Omega, \mathcal{F}, P)$ be any probability space and let $X: \Omega \rightarrow \mathbb{R}$ be any random variable. There are **three quartiles** with respect to the cumulative distribution function $F(x) = P(X \leq x)$ of the random variable X. The quartiles are:

$Q_1 = x_{0.25}$ … the first quartile or the lower quartile

… it is the quantile corresponding to the probability $p = 0.25$

$Q_2 = x_{0.5}$ … the second quartile or the median

… it is the quantile corresponding to the probability $p = 0.5$

$Q_1 = x_{0.75}$ … the third quartile or the upper quartile

… it is the quantile corresponding to the probability $p = 0.75$

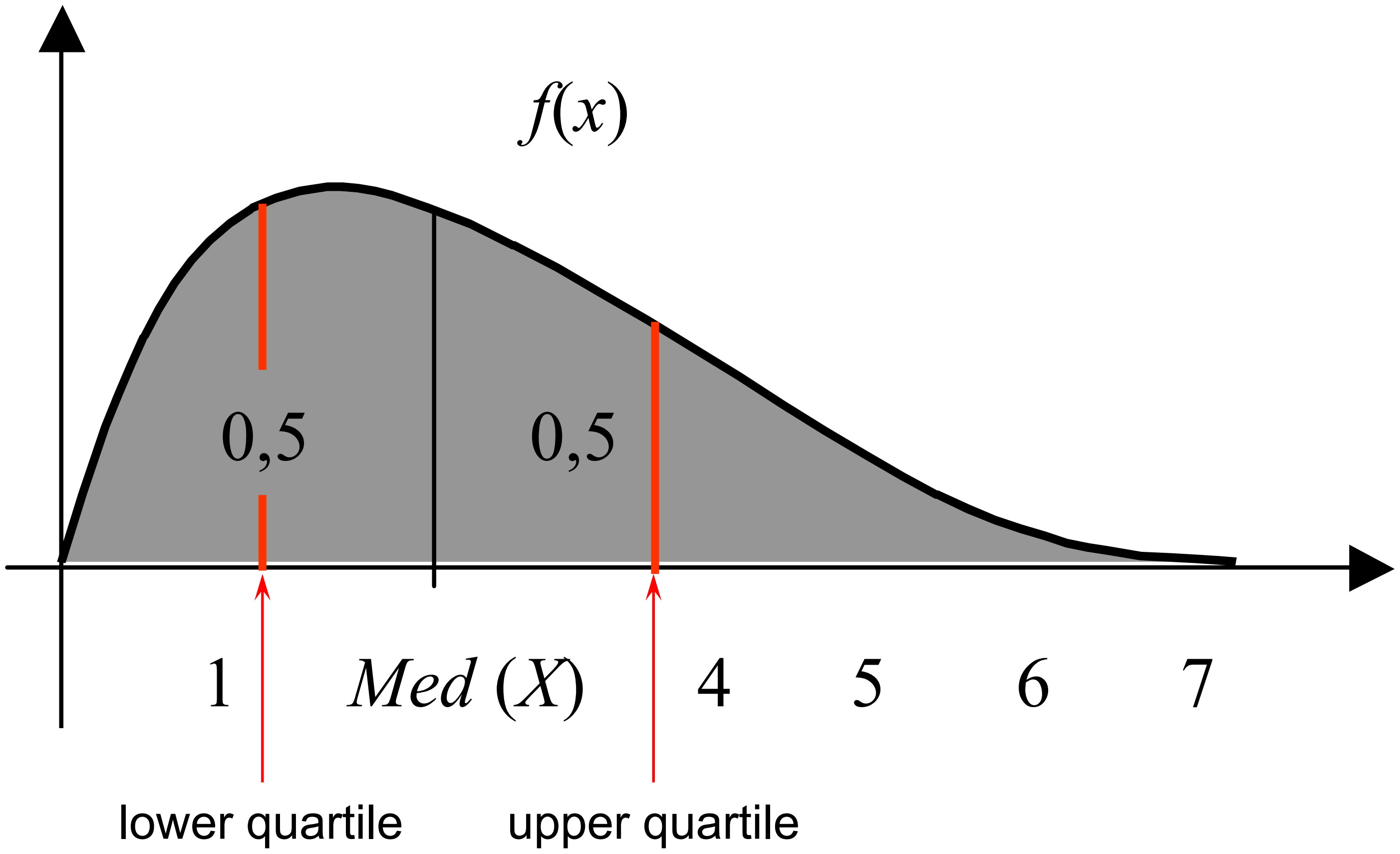
Quartiles

Notice that the difference

$$\text{IQR} = Q_3 - Q_1$$

is also called the **interquartile range**, the “midspread” or the “middle fifty”.

Median & Quartiles



Deciles

Let $(\Omega, \mathcal{F}, P)$ be any probability space and let $X: \Omega \to \mathbb{R}$ be any random variable. There are **nine deciles** with respect to the cumulative distribution function $F(x) = P(X \leq x)$ of the random variable X. The deciles are:

$$D_1 = x_{0.1} \qquad D_4 = x_{0.4} \qquad D_7 = x_{0.7}$$

$$D_2 = x_{0.2} \qquad D_5 = x_{0.5} \qquad D_8 = x_{0.8}$$

$$D_3 = x_{0.3} \qquad D_6 = x_{0.6} \qquad D_9 = x_{0.9}$$

The fifth decile $(D_5 = x_{0.5})$ is the median.

Centiles

Let $(\Omega, \mathcal{F}, P)$ be any probability space and let $X: \Omega \to \mathbb{R}$ be any random variable. There are **ninety nine centiles** with respect to the cumulative distribution function $F(x) = P(X \leq x)$ of the random variable X. The centiles are:

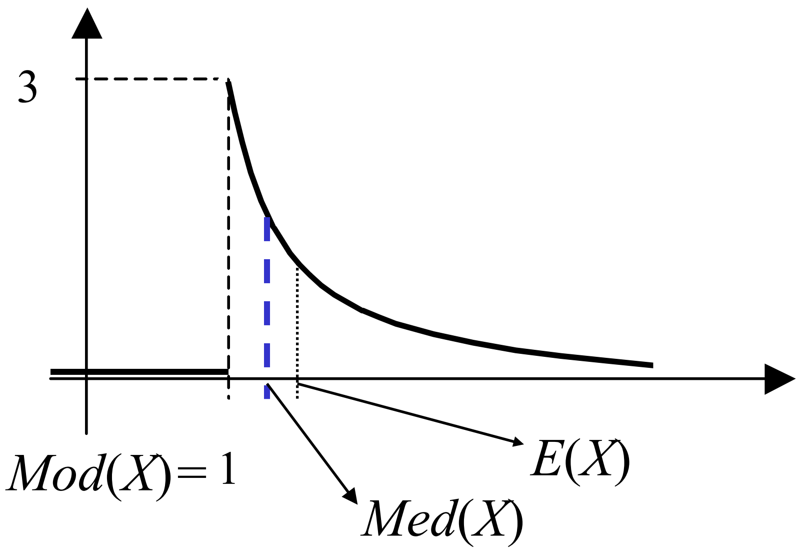
$C_1 = x_{0.01}$	...	$C_{97} = x_{0.97}$
$C_2 = x_{0.02}$	$C_{50} = x_{0.5}$	$C_{98} = x_{0.98}$
$C_3 = x_{0.03}$	...	$C_{99} = x_{0.99}$

The fiftieth centile $(C_{50} = x_{0.5})$ is the median. The twenty fifth centile and the seventy fifth centile $(C_{25} = x_{0.25}$ and $C_{75} = x_{0.75})$ is the lower quartile and the upper quartile, respectively.

Mode & Mean & Median may differ

Example: The probability density function of a random variable is

$$f(x) = \begin{cases} \frac{3}{x^4}, & x \geq 1 \\ 0, & x < 1 \end{cases}$$



Observe that $\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = \left[-\frac{1}{x^3}\right]_1^{+\infty} = 0 + 1 = 1$

Mode: $\hat{X} = 1$

Median: $\int_{-\infty}^{\tilde{X}} f(x)\,\mathrm{d}x = \left[-\frac{1}{x^3}\right]_1^{\tilde{X}} = -\frac{1}{\tilde{X}^3} + 1 = \frac{1}{2} \quad \longrightarrow \quad \tilde{X} = \sqrt[3]{2}$

Expected value: $\mathrm{E}[X] = \int_{-\infty}^{+\infty} x f(x)\,\mathrm{d}x = \left[-\frac{3}{2x^2}\right]_1^{+\infty} = 0 + \frac{3}{2} = \frac{3}{2} = 1.5$

Measures of dispersion

- Variance
- Standard deviation

Variance

Let $(\Omega, \mathcal{F}, P)$ be any probability space and let $X: \Omega \to \mathbb{R}$ be any random variable.

Assume that the expected value $\mathrm{E}[X]$ of the random variable X exists.

Notice that the sum $\mathrm{E}[X] = \sum_{\omega \in \Omega} X(\omega) p(\omega)$ (in the case II) or the integral $\mathrm{E}[X] = \int_{-\infty}^{+\infty} x f(x) \, \mathrm{d}x$ (in the case III) may diverge, i.e. either not exist at all or exist but diverge to the value $+\infty$ or $-\infty$.

Variance

Let $(\Omega, \mathcal{F}, P)$ be any probability space and let $X: \Omega \to \mathbb{R}$ be any random variable. Assuming that the expected value $\mathrm{E}[X]$ exists and is finite,

the **variance** $\mathrm{Var}(X)$ of the random variable X is

$$\mathrm{Var}(X) = \mathrm{E}[(X - \mathrm{E}[X])^2]$$

Variance

Cases I & II:

$$\sigma^2 = \mathrm{Var}(X) = \sum_{\omega \in \Omega} (X(\omega) - \mathrm{E}[X])^2 p(\omega)$$

Case III:

$$\sigma^2 = \mathrm{Var}(X) = \int_{-\infty}^{+\infty} (x - \mathrm{E}[X])^2 f(x)\,\mathrm{d}x$$

Notice that (even if we assume that $\mathrm{E}[X]$ exists and is finite), the variance may be infinite ($\mathrm{Var}(X) = +\infty$) sometimes.

Variance

Let $(\Omega, \mathcal{F}, P)$ be any probability space.
Let $X: \Omega \to \mathbb{R}$ and $Y: \Omega \to \mathbb{R}$ be two random variables.

Considering the definition of the expected value (see above) and by the properties of the sum or the integral, the following equations are easy to see under the assumption that both $\mathrm{E}[X]$ and $\mathrm{E}[Y]$ are finite:

$$\mathrm{E}[X+Y] = \mathrm{E}[X] + \mathrm{E}[Y]$$

$$\mathrm{E}[cX] = c\mathrm{E}[X] \qquad \text{for every} \quad c \in \mathbb{R}$$

Variance

The next formula holds if $\mathrm{E}[X]$ exists and is finite:

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{E}[(X-\mathrm{E}[X])^2] \\ &= \mathrm{E}[X^2 - 2X\mathrm{E}[X] + (\mathrm{E}[X])^2] \\ &= \mathrm{E}[X^2] - \mathrm{E}\big[2X\mathrm{E}[X]\big] + (\mathrm{E}[X])^2 \\ &= \mathrm{E}[X^2] - 2(\mathrm{E}[X])(\mathrm{E}[X]) + (\mathrm{E}[X])^2 \\ &= \mathrm{E}[X^2] - (\mathrm{E}[X])^2\end{aligned}$$

<u>Remark:</u> The formula may be useful in the case I to compute the variance efficiently: $\mathrm{Var}(X) = \sum_{\omega\in\Omega} X^2(\omega)p(\omega) - \left(\sum_{\omega\in\Omega} X(\omega)p(\omega)\right)^2$.

Standard deviation

The **standard deviation** is the square root of the variance:

$$\sigma = \sqrt{\mathrm{Var}(X)} = \sqrt{\sigma^2}$$

Measures of shape

- Skewness
- Kurtosis

Skewness & Kurtosis

Let $(\Omega, \mathcal{F}, P)$ be any probability space and let $X: \Omega \to \mathbb{R}$ be any random variable. Assume that the expected value $\mathrm{E}[X]$ of the random variable X exists.

Pearson's moment coefficient of <u>skewness</u> is

$$\mathrm{Skew}(X) = \mathrm{E}\left[\left(\frac{X - \mathrm{E}[X]}{\sqrt{\mathrm{Var}(X)}}\right)^3\right]$$

Pearson's moment coefficient of <u>kurtosis</u> is

$$\mathrm{Kurt}(X) = \mathrm{E}\left[\left(\frac{X - \mathrm{E}[X]}{\sqrt{\mathrm{Var}(X)}}\right)^4\right]$$

Skewness: Properties and interpretation

Pearson's moment coefficient of skewness

$$\mathrm{Skew}(X) = \mathrm{E}\left[\left(\frac{X - \mathrm{E}[X]}{\sqrt{\mathrm{Var}(X)}}\right)^3\right]$$

can be positive or zero or negative.

— $\mathrm{Skew}(X) < 0$ — the majority of the values is left to the mean

— $\mathrm{Skew}(X) = 0$ — the values are distributed ≈ symmetrically around the mean

— $\mathrm{Skew}(X) > 0$ — the majority of the values is right to the mean

Large positive or negative value — there are "outliers", i.e. values far away from the mean

Kurtosis: Properties and interpretation

Pearson's moment coefficient of kurtosis

$$\mathrm{Kurt}(X) = \mathrm{E}\left[\left(\frac{X - \mathrm{E}[X]}{\sqrt{\mathrm{Var}(X)}}\right)^4\right]$$

can be positive or zero.

— $\mathrm{Kurt}(X) \geq 0$ is small — the values are concentrated ≈ around the mean

— $\mathrm{Kurt}(X) > 0$ is large — there are "outliers", i.e.

values far away from the mean

The Skewness & Kurtosis describe the shape of the distribution of the values.

Functions of random variables

- Sample mean
- Sample variance
- Sample standard deviation

Statistic

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X_1, X_2, \dots, X_n : \Omega \to \mathbb{R}$ be random variables. A **statistic** is any function (a formula or an algebraic expression) of the random variables:

$$Y = f(X_1, X_2, \dots, X_n)$$

<u>Notice that the statistic is a new random variable.</u>

Sample mean & Sample variance

The most frequently used statistics are:

Sample mean:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Sample variance:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \bar{X})^2$$

Sample variance & Sample standard deviation

Notice that the **sample variance** satisfies the next equation:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{2n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i - X_j)^2$$

Once the sample variance s^2 is known,

the **sample standard deviation** is

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

The measures of the statistics

The sample mean:

$$\mathrm{E}[\bar{X}] = \mu$$

$$\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

The sample variance:

$$\mathrm{E}[s^2] = \sigma^2$$

The sample standard deviation:

$$\mathrm{E}[s] = \sigma$$

The expected values of the functions of random variables

- The expected value of the sample mean
- Independent events
- Independent random variables
- The variance of the sample mean
- The expected value of the sample variance

The expected value of the sample mean

Assume that the expected values $\mathrm{E}[X_1] = \mathrm{E}[X_2] = \cdots = \mathrm{E}[X_n] = \mu$.

Then

$$\mathrm{E}[\bar{X}] = \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\mathrm{E}\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu$$

Independent events

Let $(\Omega, \mathcal{F}, P)$ be a probability space.

We say that events $A, B \in \mathcal{F}$ are **independent** if and only if

$$P(A \cap B) = P(A) \times P(B)$$

so that

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A) \quad \text{and} \quad P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A)P(B)}{P(A)} = P(B)$$

$P(B) \neq 0$ $\qquad$ $P(A) \neq 0$

Independent random variables

Let $(\Omega, \mathcal{F}, P)$ be a probability space.

We say that random variables $X, Y: \Omega \to \mathbb{R}$ are **independent** if and only if

$$P(\{\omega \in \Omega : X(\omega) \leq a\} \cap \{\omega \in \Omega : Y(\omega) \leq b\}) =$$

$$= P(\{\omega \in \Omega : X(\omega) \leq a\}) \times P(\{\omega \in \Omega : Y(\omega) \leq b\}) \quad \text{for every} \quad a, b \in \mathbb{R}$$

in short:

$$P(\{X \leq a\} \cap \{Y \leq b\}) = P(\{X \leq a\}) \times P(\{Y \leq b\}) \quad \text{for every} \quad a, b \in \mathbb{R}$$

Independent random variables: Theorem

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X, Y: \Omega \to \mathbb{R}$ be independent random variables such that the expected values $\mathrm{E}[\,|X|\,]$ and $\mathrm{E}[\,|Y|\,]$ are finite.

Then

$$\mathrm{E}[X \times Y] = \mathrm{E}[X] \times \mathrm{E}[Y]$$

We prove this statement in the case I, when the sample space Ω is finite ($\Omega = \{1, 2, \ldots, N\}$). The proof uses limiting steps and some advanced results (Levi's Theorem) of the theory of measures and the Lebesgue integral.

Independent random variables: E[*XY*] = E[*X*] E[*Y*]

Proof (in the case I): Let

$$\{x_1, x_2, \dots, x_m\} = \{X(\omega) : \omega \in \Omega\} \quad \text{and} \quad \{y_1, y_2, \dots, y_n\} = \{Y(\omega) : \omega \in \Omega\}$$

be the ranges of the random variables X and Y, and let the ranges be finite.

(If the sample space Ω is finite [the case I], then so are the ranges.) Then

$$\mathrm{E}[X \times Y] = \sum_{i=1}^{m} \sum_{j=1}^{n} x_i \times y_j \times P(\{X = x_i\} \cap \{Y = y_j\})$$

Independent random variables: E[*XY*] = E[*X*] E[*Y*]

$$\mathrm{E}[X \times Y] = \sum_{i=1}^{m} \sum_{j=1}^{n} x_i \times y_j \times P(\{X = x_i\} \cap \{Y = y_j\}) =$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} x_i \times y_j \times P(\{X = x_i\}) \times P(\{Y = y_j\}) =$$

$$= \sum_{i=1}^{m} x_i \times P(\{X = x_i\}) \times \sum_{j=1}^{n} y_j \times P(\{Y = y_j\}) = \mathrm{E}[X] \times \mathrm{E}[Y]$$

Independent random variables: Theorem II

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $X', X'' : \Omega \to \mathbb{R}$ be independent random variables such that the expected values $\mu' = \mathrm{E}(|X'|)$ and $\mu'' = \mathrm{E}(|X''|)$ are finite. Then

$$\mathrm{E}[(X' - \mu')(X'' - \mu'')] = 0$$

Proof:

$$\begin{aligned}\mathrm{E}[(X' - \mu')(X'' - \mu'')] &= \mathrm{E}[X'X'' - X'\mu'' - \mu'X'' + \mu'\mu''] = \\ &= \mathrm{E}[X'X''] - \mathrm{E}[X'\mu''] - \mathrm{E}\,[\mu'X''] + \mathrm{E}[\mu'\mu''] = \\ &= \mathrm{E}[X']\mathrm{E}[X''] - \mathrm{E}[X']\mu'' - \mu'\mathrm{E}[X''] + \mu'\mu'' =\end{aligned}$$

The variance of the sample mean

Assume that the variances $\mathrm{Var}(X_1) = \mathrm{Var}(X_2) = \cdots = \mathrm{Var}(X_n) = \sigma^2$

and that the random variables $X_1, X_2, \ldots, X_n$ are pairwise independent.

Then

$$\mathrm{Var}(\bar{X}) = \mathrm{E}[(\bar{X} - \mathrm{E}[\bar{X}])^2] = \mathrm{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)^2\right] = \mathrm{E}\left[\frac{\left(\sum_{i=1}^{n}(X_i - \mu)\right)^2}{n^2}\right]$$

$$= \frac{1}{n^2}\mathrm{E}\left[\sum_{i=1}^{n}(X_i - \mu)^2 + \sum_{i=1}^{n}\sum_{\substack{j=1 \\ i \neq j}}^{n}(X_i - \mu)(X_j - \mu)\right] =$$

The variance of the sample mean

$$\mathrm{Var}(\bar{X}) = \frac{1}{n^2}\,\mathrm{E}\left[\sum_{i=1}^{n}(X_i-\mu)^2 + \sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}(X_i-\mu)(X_j-\mu)\right] =$$

$$= \frac{1}{n^2}\left(\sum_{i=1}^{n}\mathrm{E}[(X_i-\mu)^2] + \sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n}\mathrm{E}[(X_i-\mu)(X_j-\mu)]\right) =$$

The variance of the sample mean

If X_i and X_j are independent, then $\mathrm{E}\left[(X_i-\mu)(X_j-\mu)\right]=0$

$$\mathrm{Var}(\bar{X})=\frac{1}{n^2}\left(\sum_{i=1}^{n}\mathrm{E}[(X_i-\mu)^2]+\sum_{i=1}^{n}\sum_{\substack{j=1\\i\neq j}}^{n}\mathrm{E}[(X_i-\mu)(X_j-\mu)]\right)=$$

$$=\frac{1}{n^2}\left(\sum_{i=1}^{n}\mathrm{E}[(X_i-\mu)^2]\right)=$$

$$=\frac{1}{n^2}\sum_{i=1}^{n}\sigma^2=\frac{\sigma^2}{n}$$

The expected value of the sample variance

Assume that the expected values $\mathrm{E}[X_1] = \mathrm{E}[X_2] = \cdots = \mathrm{E}[X_n] = \mu$,
that the variances $\mathrm{Var}(X_1) = \mathrm{Var}(X_2) = \cdots = \mathrm{Var}(X_n) = \sigma^2$,
and that the random variables $X_1, X_2, \dots, X_n$ are pairwise independent.
Then

$$\mathrm{E}[s^2] = \mathrm{E}\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \mathrm{E}\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \frac{1}{n}\sum_{j=1}^{n}X_j\right)^2\right] =$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\mathrm{E}\left[\left(X_i - \frac{1}{n}\sum_{j=1}^{n}X_j\right)^2\right] =$$

The expected value of the sample variance

$$\mathrm{E}[s^2] = \frac{1}{n-1}\sum_{i=1}^{n} \mathrm{E}\left[\left(X_i - \frac{1}{n}\sum_{j=1}^{n} X_j\right)^2\right] =$$

$$= \frac{1}{n-1}\sum_{i=1}^{n} \mathrm{E}\left[X_i^2 - \frac{2}{n}X_i\sum_{j=1}^{n} X_j + \frac{1}{n^2}\sum_{j=1}^{n} X_j \sum_{k=1}^{n} X_k\right] =$$

$$= \frac{1}{n-1}\sum_{i=1}^{n} \mathrm{E}\left[X_i^2 - \frac{2}{n}X_i^2 - \frac{2}{n}\sum_{\substack{j=1\\ i\neq j}}^{n} X_i X_j + \frac{1}{n^2}\sum_{j=1}^{n}\sum_{\substack{k=1\\ j\neq k}}^{n} X_j X_k + \frac{1}{n^2}\sum_{j=1}^{n} X_j^2\right] =$$

The expected value of the sample variance

$$\mathrm{E}[s^2] = \frac{1}{n-1}\sum_{i=1}^{n} \mathrm{E}\left[X_i^2 - \frac{2}{n}X_i^2 - \frac{2}{n}\sum_{\substack{j=1\\ i\neq j}}^{n} X_i X_j + \frac{1}{n^2}\sum_{j=1}^{n}\sum_{\substack{k=1\\ j\neq k}}^{n} X_j X_k + \frac{1}{n^2}\sum_{j=1}^{n} X_j^2\right] =$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{n-2}{n}\mathrm{E}[X_i^2] - \frac{2}{n}\sum_{\substack{j=1\\ i\neq j}}^{n}\mathrm{E}[X_i X_j] + \frac{1}{n^2}\sum_{j=1}^{n}\sum_{\substack{k=1\\ j\neq k}}^{n}\mathrm{E}[X_j X_k] + \frac{1}{n^2}\sum_{j=1}^{n}\mathrm{E}[X_j^2]\right) =$$

The expected value of the sample variance

Recall that $\mathrm{E}[X_1] = \cdots = \mathrm{E}[X_n] = \mu$ and $\mathrm{Var}(X_1) = \cdots = \mathrm{Var}(X_n) = \sigma^2$, and $\sigma^2 = \mathrm{Var}(X_i) = \mathrm{E}[X_i^2] - (\mathrm{E}[X_i])^2 = \mathrm{E}[X_i^2] - \mu^2$ in general. Hence $\mathrm{E}[X_i^2] = \mu^2 + \sigma^2$ for every $i = 1, \dots, n$. Since X_i and X_j are independent, we have $\mathrm{E}[X_i X_j] = \mathrm{E}[X_i]\mathrm{E}[X_j] = \mu^2$ for every $i, j = 1, \dots, n$ when $i \neq j$. Therefore

$$\mathrm{E}[s^2] = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{n-2}{n}\mathrm{E}[X_i^2] - \frac{2}{n}\sum_{\substack{j=1 \\ i\neq j}}^{n}\mathrm{E}[X_i X_j] + \frac{1}{n^2}\sum_{j=1}^{n}\sum_{\substack{k=1 \\ j\neq k}}^{n}\mathrm{E}[X_j X_k] + \frac{1}{n^2}\sum_{j=1}^{n}\mathrm{E}[X_j^2]\right) =$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{n-2}{n}(\mu^2+\sigma^2) - 2\frac{n-1}{n}\mu^2 + \frac{n(n-1)}{n^2}\mu^2 + \frac{n}{n^2}(\mu^2+\sigma^2)\right) =$$

The expected value of the sample variance

$$\mathrm{E}[s^2] = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{n-2}{n}(\mu^2+\sigma^2) - 2\frac{n-1}{n}\mu^2 + \frac{n(n-1)}{n^2}\mu^2 + \frac{n}{n^2}(\mu^2+\sigma^2)\right) =$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{n-1}{n}(\mu^2+\sigma^2) - \frac{n-1}{n}\mu^2\right) =$$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{n-1}{n}\sigma^2\right) = \sigma^2$$

An alternative formula for the sample variance

We have noticed that the **sample variance** satisfies the next equation:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{2n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i - X_j)^2$$

To see the equation, note that:

$$\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n}\left(X_i - \frac{1}{n}\sum_{j=1}^{n}X_j\right)^2 = \sum_{i=1}^{n}\left(X_i^2 - \frac{2}{n}X_i\sum_{j=1}^{n}X_j + \frac{1}{n^2}\sum_{j=1}^{n}X_j\sum_{k=1}^{n}X_k\right) =$$

$$= \sum_{i=1}^{n}X_i^2 - \frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}X_iX_j + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}X_jX_k =$$

An alternative formula for the sample variance

$$\sum_{i=1}^{n}(X_i-\bar{X})^2=\sum_{i=1}^{n}X_i^2-\frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}X_iX_j+\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}X_jX_k=$$

$$=\sum_{i=1}^{n}X_i^2-\frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}X_iX_j+\frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n}X_jX_k=$$

$$=\sum_{i=1}^{n}X_i^2-\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}X_iX_j=$$

$$=\frac{1}{2n}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}X_i^2-2\sum_{i=1}^{n}\sum_{j=1}^{n}X_iX_j+\sum_{i=1}^{n}\sum_{j=1}^{n}X_j^2\right)=\frac{1}{2n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(X_i-X_j\right)^2$$